LEARNING RAW IMAGE DENOISING USING A PARAMETRIC COLOR IMAGE MODEL

Raphaël Achddou, Yann Gousseau, Saïd Ladjal

LTCI, Télécom Paris, Institut Polytechnique de Paris

ABSTRACT

Deep learning methods for image restoration have produced impressive results over recent years. Nevertheless, they generalize poorly and need large learning image datasets to be collected for each new acquisition modality. In order to avoid the building of such datasets, it has been recently proposed to develop synthetic image datasets for training image restoration methods, using scale invariant dead leaves models. While the geometry of such models can be successfully encoded with only a few parameters, the color content cannot be straightforwardly encoded. In this paper, we leverage the concept of color lines prior to build a light parametric color model relying on a chromaticity/luminance factorization. Further, we show that the corresponding synthetic dataset can be used to train neural networks for the denoising of RAW images from different camera-phones, without using any image from these devices. This shows the potential of our approach to increase the generalization capacity of learning-based denoising approaches in real case scenarios.

Index Terms— image denoising, color modeling, synthetic learning

1. INTRODUCTION

Image denoising is a crucial image restoration task. It aims at retrieving a noiseless estimate $\tilde{x}$ from a noisy observation $y = x + n$ where x is the original scene and n is a realisation of some noise, usually signal dependent. Before the advent of learning-based methods, most approaches were prior-based and often relied on some regularity assumption on images, either through variational or Bayesian methods. In the last decade, these methods were surpassed by deep learning approaches which directly aim at minimizing the reconstruction error over a large set of natural images by iteratively fitting the weights of a neural network. Such methods regularly improve performances over specific datasets but suffers from two main limitations : they need massive amounts of data to be trained efficiently and they generalize poorly to new datasets, for instance when the image acquisition modality are changed. In particular, these limitations imply heavy data collection campaigns each time a new acquisition device is considered. To circumvent this limitation, it was first proposed in [1] to build synthetic learning sets to train denoising networks, using a scale-invariant dead leaves image model. This approach led to very good restoration performances for complex restoration tasks such as RAW image denoising [2]. While these works showed that the geometrical content of natural images can be well approximated by simple models that only depend on a few parameters, the color content was encoded with heavy non-parametric representations (object are colored by directly sampling an image from a given dataset). As shown in [1], such heavy encodings of the color content cannot be replaced by simple models such as a uniform distribution over a given color space, which significantly impair the restoration performances.

In this paper, we propose a model for encoding the color content of natural images by leveraging the concept of *color line prior*, introduced in [3]. This approach enables the creation of realistic color histograms with a decoupled chromaticity/luminance representation. From this concept, we derive a color sampling algorithm that enables to build a fully parametric synthetic model to train denoising neural networks. Further, we show that by using a simple normalization procedure, a single model can successfully train networks for the RAW image denoising of different models of camera phones, without having ever seen any image originating from these phones. This shows the potential of our approach to increase the generalization capacity of learning-based denoising approaches in real case scenarios.

2. RELATED WORKS

Image restoration. The most classical approaches to image restoration tasks are based on *a priori* hypotheses about the nature of images and in particular about their mathematical regularity. These can be formulated as a regularity constraint in the minimization of an energy for variational methods (such as the total variation [4]) or in sparse-decomposition methods [5]. Non-local methods also rely on a structural a priori, the self-similarity hypothesis [6]. While these methods are reliable and have mathematical guarantees, they were often surpassed by deep learning approaches in the last decade. Advances in the design of neural networks'architectures , such as residual connections [7], or U-Net architectures [8], translated in better image denoising results.

Synthetic learning. While these learning based approaches produce impressive results, they require large amounts of data to be trained efficiently. Building such datasets is needed for each new image acquisition modality and can be challenging and time consuming. To tackle this limitation, a natural solution amounts in generating synthetic data for the training. Some recent works focus on synthesizing realistic noise with physically based models [9] or learned models [10]. Going a step further, [1] propose to synthesize both noisy and ground truth images, based on the dead leaves image model [11]. These images, which exhibit interesting statistical properties, were also shown to be efficient to learn a variety of tasks such as classification pre-training [12] or disparity map estimation [13].

Natural color distribution. Relatively few works have investigated the distribution of colors of natural images. Early works [14] have shown that a principal component analysis on natural images yields opponent-like color spaces, a hypothesis that was later theoretically investigated in [15]. Still dealing with first order statistics, empirical histograms of relatively large databases have been collected in [16]. Spatial dependency between colors have been investigated by [17] and the corresponding oscillatory patterns of opponent colors have been theoretically justified in [18]. Several color statistics have been investigated in [19] in view of a better understanding of the color constancy ability of the human visual system. In 2004, [3] introduced the color line prior, an attempt to model the distribution of colors from a single monochromatic object, which we will present in detail in the following section.

3. A PARAMETRIC COLOR IMAGE MODEL

Recall that our goal is to develop a fully synthetic model for the training of denoising network. For this, we rely on the geometry provided by scaling dead leaves model as in [2, 1] and propose a new coloring scheme completely bypassing the need of real photographs.

3.1. Background

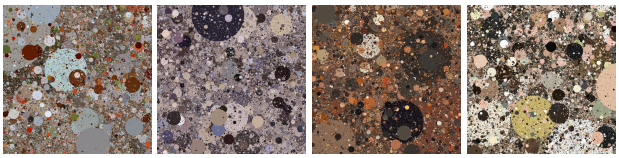


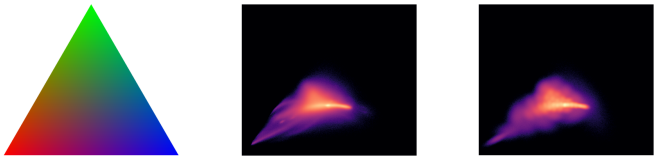
Fig. 1: Examples of dead leaves images colored with the proposed sampling scheme

Dead Leaves Model. As mentioned above, simple dead leaves images can exhibit statistical properties close to those of natural images, including the shape of the power spectrum and the distribution of the gradient [20, 21]. These images are the result of a superimposition of shapes with random color and sizes. In order to reproduce natural images statistics, it is enough to consider disks whose radii follow a power law density $f(r) = C.\mathbb{1}(r_{min} \leq r \leq r_{max}).r^{-\alpha}$ with two cut-off parameters r_{min}, r_{max}, the case $\alpha = 3$ corresponding to scale invariance. Fig. 1 displays a few example of these images.

Color Line Prior. In [3], Omer et al. show that the colors of a single Lambertian object (i.e. an object which surface diffuses light rays uniformly in all directions) are distributed along a straight line starting at $(0, 0, 0)$ in a *RAW-RGB cube*, defined as the cube where each point coordinates corresponds to the associated RGB values in the RAW image. The sensor of a camera has indeed a linear response to the number of photons. For RGB developped images, this straight lines are transformed into curves, due to the non-linearity of RAW-to-RGB transforms and tone mappings. In this work, we will consider the task of RAW images denoising and therefore adopt the color line model to generate realistic color distributions. This means that we also neglect non-lambertian objects and specularities in our model.

3.2. Color sampling algorithm

In order to sample colors to create efficient synthetic training datasets, we leverage the Color Line Prior presented in the previous section and propose to generate line clusters in the RAW domain, each one corresponding to an artificial monochromatic Lambertian object. In order to set the direction of each cluster realistically, we first study the distribution of chromaticities in natural images.



(a) Triangle of the possible directions

(b) Log-histogram of $\mathcal{P}$ in the 2D plan

(c) Log-density of the GMM fitted on $\mathcal{P}$

Fig. 2: Distribution of natural colors in the 2D representation

Estimation of the distribution of chromaticity. For this task, we use the RAISE database [22], made of RAW images from different cameras. Since the color sensitivity functions vary from one camera to another, we can not directly estimate the distribution of colors on this database. We first map the colors to a common color space, by multiplying each channel by the daylight white-balance multipliers which are estimated by camera manufacturers to fit a scene with the standard D65 illuminant. Since RAW images are organized in a Bayer frame, we consider that every (2,2) square correspond to a single R-G-G-B color. We then average the two green components to get a single value. After these oper-

ations we aggregate the RGB colors of all the pixels in the database in a large tensor $\mathcal{D}$. We then project each triplet $(r, g, b) \in \mathcal{D}$ on the plane $u + v + w = 1$ by applying the following normalization $\tilde{r}, \tilde{g}, \tilde{b} = \frac{1}{r+g+b}(r, g, b)$. to get the chromaticity values. We can visualize in Fig. 2a the space of chromaticities and in Fig. 2b the log-histogram of the colors' chromaticity for the RAISE database. In order to further reduce the dimensionality of our model, we approximate the histogram with a simple 2D Gaussian Mixture Model(GMM) of 40 components. Above 40 components, the likelihood of the model only improves marginally. Fig. 2c, illustrate this approximation. Given this parametric model, we can now sample realistic chromaticity, that define the directions of line color clusters. Next, we study the distribution of the luminance conditioned on the chromaticity to be able to sample within each color line.

Conditional luminance distribution. Here, our strategy is to analyze the distribution of the grey levels in patches of the dataset whose chromaticity is homogeneous. The underlying assumption is that patches with a homogeneous chromaticity belong to a single object. Therefore, we first extract all the disjoint (50,50) patches in the RAISE set. For each patch, we compute its average grey level $\bar{z}$, its average chromaticity $(\bar{x}, \bar{y})$, and the chromaticity's covariance $\Sigma_{x,y}$, and get rid of the patches for which $\det(\Sigma_{x,y}) \geq \eta$, where η is a chosen threshold. We noticed that the grey level mean in each bin has a heavy-tailed distribution, which we model with a Gamma distribution. We report in Fig. 3, some examples of the distribution of the grey level mean depending on the $(\bar{x}, \bar{y})$ position. In the given examples, we

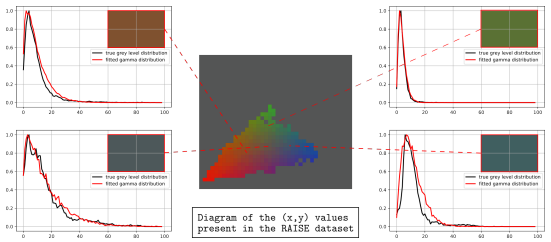


Fig. 3: Distribution of the average grey level knowing the position (x, y) in the 2D color representation.

see that the parameters of the Gamma distribution depend on $(\bar{x}, \bar{y})$. As a simple approximation, we consider that the distribution of the average grey level can be expressed as : $z|(x, y) \sim \lambda(x, y)\Gamma(\theta_{\text{grey}})$, where θ_{grey} are the Gamma distribution's parameters found for the central bin of the triangle, which corresponds to grey colors. The value $\lambda(x, y)$ corresponds to the maximal grey level attainable for the chromaticity (x, y). Given (x, y), we can easily retrieve a color triplet $(r, g, b)_{x,y}$ in the plane defined by $u + v + w = 1$. Further, the value $\lambda(x, y)$ is then empirically defined by the following formula : $\lambda(x, y) = (3.\max((r, g, b)_{x,y}))^{-1}$.

Sampling algorithm. We can now easily derive a point cloud generation algorithm for a single object, thanks to the models for $p(x, y)$ and $p(\bar{z}|x, y)$. We begin by sampling an average chromaticity $\mu = (\bar{x}, \bar{y})$ using the 2D GMM. However, we observed that using a single chromaticity is not enough to model variations within an object. Therefore, we sample each point's chromaticity following a 2D-Gaussian distribution centered on μ. This sampling yields a set $C = \left\{(x_i, y_i) \sim \mathcal{N}(\mu, \sigma^2.I)\right\}_{i\in[1,\dots,N]}$, where N is the desired number of points. The standard deviation σ follows a uniform distribution $\sigma \sim \mathcal{U}([5.e^{-5}, 5.e^{-4}])$, to increase diversity. We can now sample the luminance average $\bar{z}$ conditioned on μ, with the Gamma law defined above. Many parametric grey-level distribution models are possible given $\bar{z}$. For simplicity, we propose to simply model it by a Gaussian distribution $\mathcal{N}(\bar{z}, \beta^2)$, where $\beta \sim \mathcal{U}([0.05, 0.3])$.

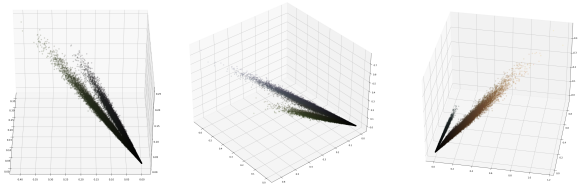


Fig. 4: 3 different color points cloud for dead leaves image generation

The described algorithm allows us to sample colors for a single monochromatic object by following the Color Line Prior. Now, realistic images contain several objects. In order to include realistic objects transition within the generated synthetic dataset, we create point clouds from two different color clusters. In Fig. 4, we give examples of the obtained point clouds. In order to also mimic monochromatic textures, we chose to synthesize another set of dead leaves images with only one color cluster. The idea here is that this simple color sampling method is sufficient to model the interactions between objects and monochromatic textures.

4. RAW IMAGE DENOISING EXPERIMENTS

In this section, we report results of RAW image denoising based on the SIDD dataset [23], which contains noisy RAW images from 5 different cameraphones. It is worth mentioning that images captured by these cameras were never used during training nor to estimate the distribution of color.

Pre-processing. As explained before, we wish to model the color distribution of digital photographs from a universal color representation space. Nevertheless, each specific camera has its own color specifications and we need to invert the daylight white balance step with camera-specific parameters. Since these parameters are not available in the SIDD database, we use random values sampled around realistic average estimates, based on the white balance parameters available in the dataset. This would be problematic to develop real photographs, but enables us to generate training

Table 1: Numerical evaluation of the denoising of smartphone RAW images from the SIDD database [23]. For each training database we report the PSNR per camera (IP:Iphone 7, GP: Google Pixel, S6: Samsung S6, N6:Nexus 6, G4: LG G4) in the RAW and sRGB domain, as well as on the whole test set.

Training database	PSNR RAW					
	IP	GP	S6	N6	G4	Global
SIDD-DL	**54.82**	**48.33**	**44.92**	**48.09**	**50.30**	**49.94**
2DM-DL	**54.60**	**48.18**	**44.18**	**47.95**	50.09	**49.63**
3DM-DL	53.91	47.94	43.99	47.78	**50.24**	49.31
RAISE (not norm.)	53.74	48.00	44.18	47.63	49.84	49.26
(RAISE-DL)	(54.59)	(48.24)	(44.21)	(47.95)	(50.40)	(49.68)

PSNR RGB					
IP	GP	S6	N6	G4	Global
40.33	**37.70**	**35.43**	**36.13**	**37.34**	**37.84**
39.66	**37.51**	**34.43**	**35.63**	**36.76**	**37.32**
39.54	37.27	34.34	35.51	36.68	37.14
38.83	36.86	34.36	35.01	36.10	36.65
(40.28)	(37.64)	(34.67)	(35.54)	(37.26)	(37.56)

sets with a wide enough variety of realistic colors. Specifically, we chose to model the red and blue gain with uniform distributions, $g_r \sim \mathcal{U}([1.9, 2.4])$ and $g_b \sim \mathcal{U}([1.5, 1.9])$, and to reverse the white balance operation as it is done by Brooks et al. [24].

Synthetic datasets. To evaluate our approach, we generate different synthetic sets. First, we follow the method presented in [2] and sample the colors directly from the SIDD training set (SIDD-DL). This dataset has its geometry given by the deal leaves model and its colors sampled from images acquired with the same cameras as those used for the images that we will denoise (SIDD test set). As relatively simple baseline for a synthetic dataset not using the SIDD color distribution, we approximate the color distribution in the RGB cube with a 3D-GMM estimated on the RAISE dataset (3DM-DL). Eventually, we generate a set with our approach (2DM-DL), with the parametric color distribution described in the previous section and estimated on the RAISE dataset. In order to assess the limitation caused by the discrepancy between the color distributions of the RAISE and SIDD dataset, we also generate dead leaves images with the colors sampled directly from RAISE images (RAISE-DL) and processed as explained above.

Network and Training. Each dataset is composed of 50k (256,256) RAW patches. We model RAW noise as a Poisson-Gaussian noise, which parameters were estimated in [2] on the SIDD dataset. We train a UNet network to denoise RAW images in a residual manner, inspired by [7]. We minimize the $\mathcal{L}_1$ loss for 500 epochs with a simple learning rate decay starting from $1e^{-4}$ to $1e^{-6}$, with a fixed batch size of 64.

Denoising Results. As reported in Table 1, we see that our approach (2DM-DL) performs close to SIDD-DL, without ever having access to the original set. We also note that our methods is surprisingly close to the RAISE-DL approach which stands as an upper bound of our performances (less than 0.1 dB difference). This indicates that the color line model is a sound model to approximate histograms of natural images, and that the performance gap is mainly caused by the discrepancy between the color distribution of the two datasets. Visually, we observe in Fig. 5 that the denoising performances of 2DM-DL is almost as good as good as SIDD-DL and retrieves much better the details than 3DM-DL, confirming the intuition that an over-simplistic color model leads to poor performances. Finally, we also show that directly training with RAW images from the RAISE set without any color processing leads to poor performances, even though the geometry is more accurate than with dead leaves images.

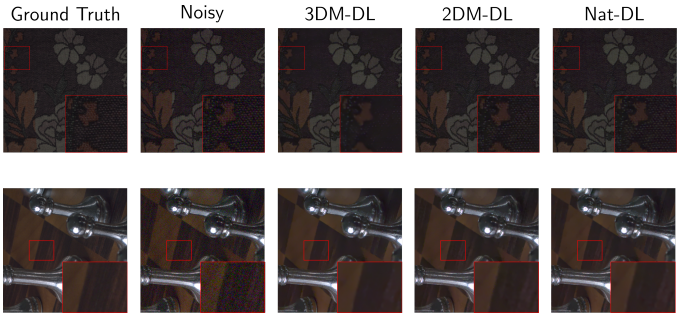


Fig. 5: Comparison of the different denoising models. From left to right : ground truth image, noisy image, 3DM-DL denoising,2DM-DL denoising,Nat-DL denoising.

5. CONCLUSION

In the present paper we introduced a novel parametric color sampling algorithm that relies on a relatively small number of parameters. This algorithm allows one to generate realistic color point clouds, which, when combined with the geometry of dead leaves model, can in turn be used to train a RAW image denoiser efficiently. In particular, this synthetic training set is obtained without using any information from the training set of the considered SIDD dataset. The obtained performances are very close to the ones obtained by directly sampling colors in the training set. In future works, we hope to reduce this gap by investigating the transition from one RAW image dataset to another, as it seems to be the main cause for the slight loss of performances.

6. REFERENCES

[1] Raphaël Achddou, Yann Gousseau, and Saïd Ladjal, "Synthetic images as a regularity prior for image restoration neural networks," in *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer, 2021, pp. 333–345.

[2] Raphaël Achddou, Yann Gousseau, and Saïd Ladjal,

"Fully synthetic training for image restoration tasks," preprint HAL, Jan. 2023.

[3] Ido Omer and Michael Werman, "Color lines: Image specific color representation," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.* IEEE, 2004.

[4] Leonid I Rudin, Stanley Osher, and Emad Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[5] David L Donoho and Iain M Johnstone, "Minimax estimation via wavelet shrinkage," *The annals of Statistics*, vol. 26, no. 3, pp. 879–921, 1998.

[6] Antoni Buades, Bartomeu Coll, and J-M Morel, "A non-local algorithm for image denoising," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05).* IEEE, 2005, vol. 2, pp. 60–65.

[7] Kai Zhang, Wangmeng Zuo, and Lei Zhang, "Ffdnet: Toward a fast and flexible solution for cnn-based image denoising," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4608–4622, 2018.

[8] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[9] Kaixuan Wei, Ying Fu, Yinqiang Zheng, and Jiaolong Yang, "Physics-based noise modeling for extreme low-light photography," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[10] Abdelrahman Abdelhamed, Marcus A Brubaker, and Michael S Brown, "Noise flow: Noise modeling with conditional normalizing flows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 3165–3173.

[11] George Matheron, "Random sets and integral geometry," 1975.

[12] Manel Baradad, Jonas Wulff, Tongzhou Wang, Phillip Isola, and Antonio Torralba, "Learning to see by looking at noise," *Advances in Neural Information Processing Systems*, vol. 34, 2021.

[13] Pavan C Madhusudana, Seok-Jun Lee, and Hamid Rahim Sheikh, "Revisiting dead leaves model: Training with synthetic data," *IEEE Signal Processing Letters*, 2021.

[14] Yu-Ichi Ohta, Takeo Kanade, and Toshiyuki Sakai, "Color information for region segmentation," *Computer graphics and image processing*, vol. 13, no. 3, pp. 222–241, 1980.

[15] Gershon Buchsbaum and Allan Gottschalk, "Trichromacy, opponent colours coding and optimum colour information transmission in the retina," *Proceedings of the Royal society of London. Series B. Biological sciences*, vol. 220, no. 1218, pp. 89–113, 1983.

[16] Baptiste Mazin, *Méthodes robustes pour l'estimation d'illuminants et la prise en compte de la couleur en comparaison d'images*, Ph.D. thesis, Télécom ParisTech, 2014.

[17] Daniel L Ruderman, Thomas W Cronin, and Chuan-Chin Chiao, "Statistics of cone responses to natural images: implications for visual coding," *JOSA A*, vol. 15, no. 8, pp. 2036–2045, 1998.

[18] Edoardo Provenzi, Julie Delon, Yann Gousseau, and Baptiste Mazin, "On the second order spatiochromatic structure of natural images," *Vision research*, vol. 120, pp. 22–38, 2016.

[19] Jürgen Golz and Donald IA MacLeod, "Influence of scene statistics on colour constancy," *Nature*, vol. 415, no. 6872, pp. 637–640, 2002.

[20] Luis Alvarez, Yann Gousseau, and Jean-Michel Morel, "The size of objects in natural and artificial images," in *Advances in Imaging and Electron Physics*, vol. 111, pp. 167–242. Elsevier, 1999.

[21] Ann B Lee, David Mumford, and Jinggang Huang, "Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model," *International Journal of Computer Vision*, 2001.

[22] Duc-Tien Dang-Nguyen, Cecilia Pasquini, Valentina Conotter, and Giulia Boato, "Raise: A raw images dataset for digital image forensics," in *Proceedings of the 6th ACM multimedia systems conference*, 2015, pp. 219–224.

[23] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown, "A high-quality denoising dataset for smartphone cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1692–1700.

[24] Tim Brooks, Ben Mildenhall, Tianfan Xue, Jiawen Chen, Dillon Sharlet, and Jonathan T Barron, "Unprocessing images for learned raw denoising," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11036–11045.